HUMANVERIFIED
CHAIN OF CUSTODY

The Evidence Trail

HUMANVERIFIED CHAIN OF CUSTODY

Concept paper and proof-of-concept invitation for clearer, bounded records of how work moves from source to outcome.

> ***A clearer record of how work moved from source to outcome.***

STATUS	Proposed proof of concept
VERSION	v0.1
DOCUMENT ID	HV-COC-WP-2026-09-03-v0.1
PUBLICATION DATE	3 September 2026
PREPARED FOR	public readers, prospective collaborators and funders

This publication describes a proposed engineering evidence trail. It is not a deployed verification service, a legal certificate, or proof of complete authorship or truth.

Executive summary

Work now moves through more hands, tools, edits and handovers than its final form usually reveals. A report may draw on a source set, several drafts and a code build. A film may pass through capture, edit, review and export. A crafted object may move from material inputs through making, inspection and transfer. The finished result can be useful and persuasive while leaving reasonable questions unanswered: what was supplied, what changed, who reviewed a stated step, and where does the available evidence stop?

HumanVerified Chain of Custody is a proposal to test a practical response. It would create a HumanVerified Evidence Record: a structured, versioned account of selected events connecting identified inputs, participants, tools and outputs. Its readable view is the Evidence Trail. The record is intended to make evidence and limits visible together, rather than issue a broad badge of authenticity or character.

The proposed proof of concept will test one event and verification model across three deliberately different kinds of work:

1. research, documentation and code;
1. moving image and editorial production; and
1. manual craft and physical custody.

The question is not whether every claim about a work can be made certain. It is whether a proportionate record can help a reader distinguish a participant declaration from machine-verifiable evidence, a documented review, and an unresolved gap. That distinction is useful to creators, commissioners, collaborators and recipients of work. It may also expose where a record adds burden without adding enough value.

This is a proposal, not a live service or a claim of results. The PoC is designed to produce evidence about the method itself: which records people understand, what controls resist realistic misuse, what information should remain private, and where the system must decline to make a wider claim.

1. The gap between a claim and a readable record

Most provenance language begins at the end. An item is described as handmade. A video is presented as an account of an event. A document is attributed to an organisation. A software release is said to come from a repository. Those statements may be accurate, incomplete, contested or unsupported. The difficulty is not that readers are incapable of judgement; it is that they often receive too little structured information to exercise it.

Existing tools solve parts of the problem. File hashes can identify exact bytes. Version control can record a source history. Signed build attestations can connect a build to declared inputs. Content Credentials can carry assertions with media. Physical labels can link an object to a record. None of these, alone, explains the whole journey or turns an assertion into a fact about the real world.

The proposed programme begins with a narrower aim: record selected, policy-defined events and make the nature of each assertion readable. A recipient should be able to see what the record supports, what it does not establish, how recently it was checked, and whether it has since been superseded, withdrawn or cannot be verified.

HumanVerified already uses this discipline in two different settings. Marketplace records can show provenance and custody evidence without calling an object universally authentic. An Assurance Reference can link a vetted account holder to a defined interaction without becoming a character reference. Chain of Custody extends the same evidence-with-boundaries approach; it does not merge those products.

HumanVerified term	Bounded question
HumanVerified Assurance Reference	Is the participant in this interaction linked to an account holder vetted to the displayed scope?
HumanVerified Evidence Record	What declared, evidenced, reviewed and unresolved events connect specified inputs, participants, tools and artifact versions?
Evidence Trail	What is the readable public summary of that record and its limits?

2. The proposed response

An Evidence Record would attach a purpose, scope and version history to a defined work or item. It could record a participant declaration, an artifact digest, an input or output version, a review event, an allowed public disclosure, and a current status. It would not expose all underlying material by default. Public views should disclose only privacy-reviewed receipts and status information; private evidence should remain access-controlled and retention-limited.

The proposed model has four visible evidence layers. They must remain distinct even when they appear in the same timeline.

1. Declared — a participant has stated an event, action, tool use, contribution or limitation and bound that statement to their checkpoint credential.
1. Evidenced — machine-verifiable data, or an independently recorded observation under the displayed policy, supports a stated event. Exact artifact bytes may match a recorded digest; a camera segment or log receipt may be present.
1. Reviewed — a named reviewer has assessed material to a displayed scope. Review records what the reviewer considered and the decision made; it does not silently expand the underlying evidence.
1. Unresolved — evidence is missing, unavailable, disputed, omitted, outside scope or insufficient to support the proposed assertion.

Cryptographic validation and human review are separate dimensions. A valid signature does not make a claim true. A reviewer’s assessment does not convert unobserved events into evidence. The point of the system is not to conceal that separation behind a green badge.

What an Evidence Record can support

It can support a statement that…	It cannot establish by itself…
exact bytes match a recorded digest	semantic truth, or that a recorded scene genuinely occurred as represented
a registered credential completed a signed checkpoint under the stated policy	complete human authorship, or the absence of undeclared tool use
a server accepted and logged a named event at a recorded time	ownership, originality, copyright, consent or lawful use
an artifact version names specified parent or ingredient versions	competence, character, safety, quality or suitability
an assertion, review and known limitation were recorded to a stated scope	that every action was observed, declared or included
a receipt is included in a witnessed log checkpoint	that a physical tag remained attached to one particular object
the current status is active, superseded, revoked, expired or unavailable	legal chain-of-custody status or admissibility

The absence of a declared AI contribution does not establish tool absence. It means only that no such contribution was recorded for the stated scope.

3. Three PoC tracks

The PoC should not begin by trying to make one claim fit every kind of work. It should test whether a common event model is useful across three tracks with different failure modes, privacy questions and reader expectations.

Research, documentation and code

This track will record selected sources, research snapshots, declarations of relevant tool use, drafts, review checkpoints, source commits, build inputs and released outputs. It can demonstrate how a reader moves from a final publication or build to the particular versions and statements named by the record.

Git can identify source versions, but it does not establish the identity or intention of a natural person. in-toto and SLSA-style attestations can bind a declared build step and its subjects to a signed statement, but they are build-provenance mechanisms rather than signals of human authorship. The PoC should make that boundary clear in both record data and reader language.

This concept paper is itself a pre-PoC example of transparent tool disclosure. It is not evidence that the proposed protocol has already been implemented or validated.

Moving image and editorial production

This track will test a journey from brief and supplied material through capture, editing, review and final export. The proposed capture exercise uses a primary camera and a witness camera, with separate keys and, where practicable, distinct administrative or failure domains. Both receive fresh server challenges during a session. A visible and audible response to the challenge is cross-recorded, and short closed segments are hashed and sequenced.

The challenge raises the cost of simply replaying a prepared sequence and ties two streams to a fresh session. It does not make replay impossible. A high-resolution display, optical replay, feed injection, compromised capture path, virtual camera or coordinated participant can still defeat or confuse the evidence. Those are test cases, not hidden exceptions.

For editing, the proposed record captures ingredient hashes, project and configuration hashes, declared edit and generative actions, application and plugin versions, export settings, reviewer decisions and handovers. Where a target format supports it, C2PA Content Credentials can bind declared ingredients and assertions to a media asset. A separate HumanVerified receipt and status check would show the policy-specific record. Neither mechanism makes the underlying assertions true merely because the structure validates.

Manual craft and physical custody

This track will test how material inputs, work-in-progress checkpoints, maker declarations, reviewer observations, custody transfers and final inspection can be recorded for a physical object. The record should support an honest account of what was seen and when, without claiming to observe every moment of making.

The experiment may test a tamper-evident label and a cryptographic challenge-response NFC tag. A QR code or NFC identifier alone is readily copied. Even a cryptographic tag shows possession of the responding tag; it does not prove continuous attachment to the same object. The tests should deliberately attempt tag transfer, substitution, staged work, display replay and omitted steps. A control that fails to detect an attack is an evidence boundary to document, not a result to smooth away.

4. A fictional Evidence Record

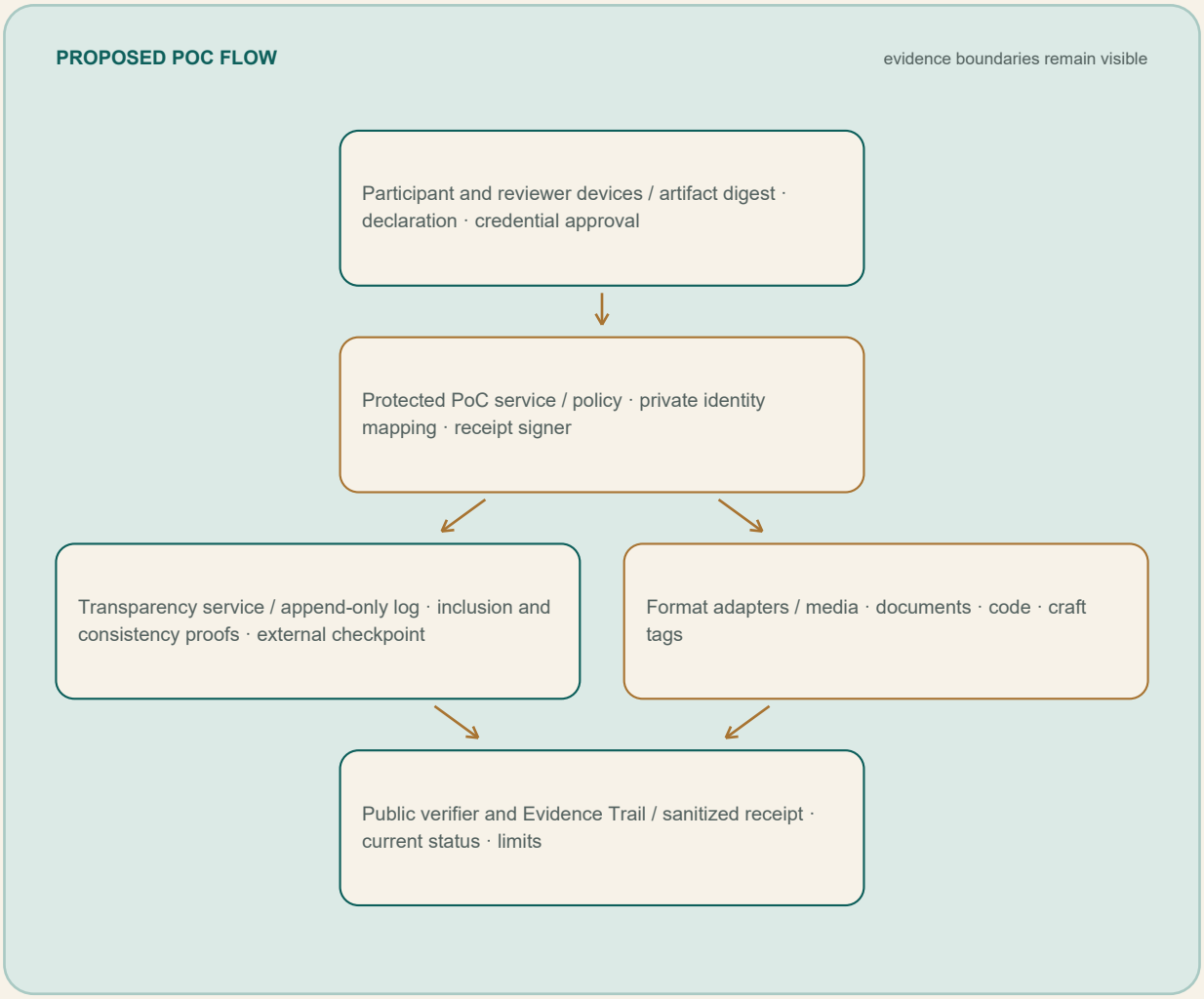
The following is illustrative only. It is not a live record, a real person, a real artifact, or a verification claim.

```
HumanVerified Evidence Record: HV-ER-DEMO-6F7K-29Q
Status: Illustrative only
Purpose: Demonstrate a bounded craft-to-publication evidence trail
Artifact: Demo hand-finished notebook, version 3
Public scope: selected making, review and handover events

2026-09-03 09:14  DECLARED
Maker states that recycled paper and linen thread were used.
Bound to a participant checkpoint. Material origin not independently checked.

2026-09-03 11:42  EVIDENCED
Closed image segment digest recorded; witness-camera segment references
the same fresh session challenge. The images support the recorded session,
not continuous observation of every making step.

2026-09-03 14:20  REVIEWED
Reviewer inspected the named version against the displayed material and
image record. Review scope: visible construction and record consistency.

2026-09-03 14:32  UNRESOLVED
No independent evidence was collected for paper supply-chain claims.

Current status: illustrative only; no live verification route
```

The public verifier envisaged by the PoC would present this sort of record as a timeline, alongside artifact digest matching, receipt status, policy version, reader-facing limits and a route to report a problem. It should not rely on a static screenshot. A screenshot can be copied after a status changes; a verifier needs to resolve the current record.

5. Technical outline: evidence, not a single trust root

The proposed service separates four trust zones:



In the participant zone, a device can hash a closed artifact, make a declaration and approve a fresh challenge through a participant-bound credential. The baseline proposal uses WebAuthn user verification for named checkpoints. It indicates that the registered credential responded for the relevant relying party under the protocol; it does not establish civil identity, exclusive control, informed intent or authorship. Higher-tier hardware and attestation evidence must be displayed as a tier, not assumed universally.

In the protected service zone, policy validation, private identity mapping, evidence retention and receipt signing occur separately from the public website. A live pilot would need protected service keys, a secure key-management design, access controls, correction and revocation workflows. The current HumanVerified sites are publishing surfaces, not this service.

In the transparency zone, accepted receipts enter an append-only Merkle log with signed tree heads, inclusion proofs and consistency proofs. Selected checkpoints should be independently witnessed or timestamped. An ordinary database labelled “append-only” is not enough: it cannot by itself make later operator equivocation or deletion detectable. Hash chains similarly detect changes only relative to a trusted anchor; they do not create trust in isolation.

Steam is not the trust root for this system. It may be explored later as an optional distribution or account-authentication channel. Its DRM, library controls and anti-cheat services do not validate camera photons, browser truth, capture inputs, editor completeness or legal identity. Treating a platform account as

proof of those matters would create a misleading record.

6. Testing and evaluation

The PoC hypothesis is modest: a shared, bounded event model may make selected evidence easier to interpret without imposing disproportionate capture, privacy or review burden. That is a hypothesis to test, not a performance forecast.

The programme should agree measures before trials and publish the methods used. Suggested measures include:

1. reader comprehension of what a record supports and does not establish;
1. time and effort required for each participant and reviewer checkpoint;
1. rate and reason for missing, disputed or unresolved evidence;
1. success of correction, supersession, withdrawal and revocation handling;
1. accessibility findings in the record and verifier journeys;
1. attack-detection coverage for replay, substitution, omitted actions, tag transfer and log inconsistency; and
1. privacy and data-minimisation findings, including fields that were collected but did not prove useful.

The tests should include adversarial attempts rather than only happy-path demonstrations. Results should report missed attacks, misunderstanding and operational friction as well as successful validation. A record that invites over-reliance has failed even if its cryptography operates as designed.

Before any human trial, the operator must complete privacy and security design work appropriate to the actual pilot: a data map, DPIA screening, lawful-basis and retention analysis, processor review, rights workflow, incident plan, accessibility assessment and an exclusion route for people who cannot or should not take part. Video, audio and device data require particular care. Candidate or worker participation must not become covert or compulsory monitoring, and missing evidence must not determine hiring suitability.

7. Non-goals and decision boundaries

The programme is not proposing a universal authenticity badge, a social credit score, a general identity credential, a background-check substitute, or an automated hiring tool. It is not intended to settle copyright, ownership, consent, legal admissibility, product safety or the truth of a scene. It is also not designed to expose private prompts, raw evidence, identity documents, stable device identifiers or private locations through a public ledger.

An Evidence Record should be purpose-bound and proportionate. A craft record should not become an identity dossier. A candidate-work-sample trail should not become surveillance or an automated decision. A media receipt should not be treated as a statement that every frame is truthful. The more consequential the decision, the more important it is to name the separate checks and human judgement that remain necessary.

8. The case for funders and collaborators

The immediate opportunity is a disciplined experiment, not a claim that provenance has been solved. A relatively small PoC can compare one common record model across digital, editorial and physical work, then show precisely where the approach is useful, costly, confusing or unsafe to scale.

For funders, the value lies in reducing a common category error: treating an opaque final output as if it carried enough context for high-confidence reliance. The programme tests an alternative in which the record makes both evidence and uncertainty visible. Its outputs should be reusable research: event fixtures, fictional

examples, policy language, verifier patterns, threat findings, accessibility observations and a publishable evidence-gap report.

For collaborators, the ask is bounded. Technical partners can test credential, transparency, media and format-adapter assumptions. Editorial and production partners can test whether the record fits real working practice. Makers and custody partners can test physical handovers without being asked to perform certainty they do not possess. Privacy, identity-assurance, security and accessibility specialists can challenge the design before it reaches a live pilot.

The next step is to decide whether the PoC can be run responsibly with a defined cohort, clear governance and measurable questions. If it cannot, the responsible outcome is to narrow or stop the proposal—not to publish a stronger promise.

9. Numbered assumptions

1. A controlled six-week PoC can be staffed, funded and governed before participant work begins. If not, the scope and timetable must change.
2. Two capture devices can be operated with meaningfully distinct keys and failure domains. If not, their corroboration value is lower than proposed.
3. A selected C2PA implementation can create records for the formats chosen for the media track. Otherwise the PoC must use detached manifests or a narrower format set.
4. Available browsers and authenticators will expose the WebAuthn user-verification and attestation behaviour required by the selected evidence tiers. The technical lead must test the actual fleet.
5. Appropriate timestamp, independent witness, protected-key and signing suppliers can be selected within the pilot’s cost and governance constraints.
6. Participants will be consenting adults in a controlled cohort. Different cohorts may require further safeguarding, employment and legal design.
7. Public opaque identifiers can be designed and tested to resist practical correlation. If they cannot, public disclosure must be reduced.
8. The current HumanVerified websites will remain static publication surfaces, with any live evidence service designed and operated separately.
9. This v0.1 document has not been independently human-reviewed or approved by a named reviewer. It must not be represented as such.

10. Source notes

The sources below are primary standards or official guidance used to frame the proposal. They are not endorsements of HumanVerified, and no conformance or certification is claimed. All links accessed 3 September 2026.

1. Coalition for Content Provenance and Authenticity, C2PA Technical Specification 2.4: https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html
2. Coalition for Content Provenance and Authenticity, Conformance Program: https://c2pa.org/conformance/
3. W3C, Web Authentication: An API for accessing Public Key Credentials, Level 3: https://www.w3.org/TR/webauthn-3/
4. IETF, RFC 8785: JSON Canonicalization Scheme: https://www.rfc-editor.org/rfc/rfc8785

5 IETF, RFC 3161: Internet X.509 Public Key Infrastructure Time-Stamp Protocol: https://www.rfc-editor.org/rfc/rfc3161

6 IETF, RFC 9162: Certificate Transparency Version 2.0: https://www.rfc-editor.org/rfc/rfc9162

7 in-toto, Statement v1 specification: https://github.com/in-toto/attestation/blob/main/spec/v1/statement.md

8 Supply-chain Levels for Software Artifacts, Provenance v1.2: https://slsa.dev/spec/v1.2/provenance

9 NIST, SP 800-63B-4: Digital Identity Guidelines—Authentication and Authenticator Management: https://pages.nist.gov/800-63-4/sp800-63b/authenticators/

10 Information Commissioner’s Office, Data minimisation: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/data-minimisation/

11 Information Commissioner’s Office, Pseudonymisation: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/

12 UK Government, UK digital verification services trust framework collection: https://www.gov.uk/government/collections/uk-digital-verification-services-trust-framework

Annex A — AI generation and use disclosure

A.1 Status

This document was generated using OpenAI Codex under human direction for everwished innovation on 3 September 2026. The AI-generated material includes this annex. “Responsible-AI disclosure” describes the disclosure practice; it is not an external certification or assurance mark.

This document describes a proposed proof of concept. Its publication is not evidence that the proposed system has been implemented, independently audited, or shown to achieve any stated outcome.

A.2 Human inputs and direction

Human input supplied the problem statement, intended audiences, approved technical thesis, delivery constraints, existing HumanVerified materials, and required claim boundaries. Human direction is not the same as independent human validation or approval.

A.3 AI contribution

AI systems assisted with research retrieval, architecture development, threat analysis, drafting, editing, diagram specification, consistency review, document generation, and release-quality checks. The source list identifies the external materials relied upon. Tool or model identifiers are stated only where they were captured accurately; none has been inferred.

A.4 Review state

At publication, no claim should be described as human-reviewed unless a named reviewer has completed and recorded that review. AI-based cross-checking is not independent human review. Legal, privacy, identity-assurance, cryptographic, accessibility, and operational decisions require qualified human assessment before a live pilot.

A.5 Fact and assumption controls

Externally verifiable claims are cited to primary sources where available. Assumptions, proposals, untested hypotheses, and limitations are identified as such. Citations show the basis for a statement; they do not guarantee that the statement is complete or suitable for a particular implementation.

A.6 Known limitations

Generative AI can omit context, misstate sources, produce plausible but incorrect details, and conceal uncertainty in fluent language. The proposed evidence system would provide bounded provenance signals, not semantic truth, guaranteed authorship, absence of AI contribution, legal ownership, safety, or resistance to every alteration. Cryptographic validation establishes only the properties defined by the relevant keys, records, trust anchors, and verification policy.

A.7 Document provenance and integrity

The editable Markdown is the publication source. Each PDF contains a visible document identifier, page count, annex marker, and source-hash prefix. The external release manifest records the complete source and final-PDF hashes. PDF permission controls deter editing in conforming software but can be bypassed by other tools. A changed copy is detectable only when compared with a trusted checksum or signature.

This release is not claimed as PDF/UA conformant. It provides selectable text, embedded fonts, `en-GB` language metadata and permission for accessibility extraction, but it is not a fully tagged PDF. Tagged-PDF remediation and independent accessibility review remain required before a live publication decision.

A.8 Privacy

The publication records input categories and source references rather than a private prompt transcript. It must not contain identity documents, raw participant evidence, secrets, credentials, device identifiers, private locations, or signing material. Public ledgers should contain only privacy-reviewed opaque identifiers and cryptographic commitments; sensitive evidence belongs in access-controlled, retention-limited storage.

A.9 Change control

This package is an unpublished release candidate until the owner deliberately publishes it and its checksum manifest through a trusted channel. Candidate rebuilds in this workspace may replace local v0.1 files and necessarily create new hashes; the builder requires an explicit release-candidate replacement flag when prior output exists.

After a package has been distributed, cited or otherwise relied upon, any byte change must use a new version or build identifier and preserve the earlier package and manifest. Publication must use a versioned, package-atomic deployment or equivalent rollback-controlled process. The local document builder uses atomic replacement per file but does not claim package-level atomic deployment.